MODELES DE DUREE DE VIE

Cours 1 : Introduction

I- Contexte et définitions

II- Les données

III- Caractéristiques d'intérêt

IV- Evènements non renouvelables/renouvelables (unique/répété)

I- Contexte et définitions

Définition : On appelle *durée de vie* une variable aléatoire T positive, généralement, la durée s'écoulant entre deux *évènements*.

Ex d'évènements : mort, panne, sinistre, entrée en chômage, maladie.

Modèles de durée de vie : sous-ensemble de méthodes statistiques adaptées à l'étude des durées de vie.

Objectifs : Modélisation, en fonction de facteurs explicatifs éventuels

- ✓ de la loi de la durée T (évènement non renouvelable) : instant d'occurrence de la mort, de la premiere, panne, de la ruine,..
- ✓ De la loi du processus temporel de durées $(T_1, \ldots . T_n)_{n \in N}$ (évènement renouvelable) : instants d'occurrences de pannes, de sinistres, de rechutes successifs..

I- Contexte et définitions

Champ d'application des modèles de durées :

- ✓ Fiabilité : durée de la vie d'un matériel, durée entre deux pannes d'un matériel réparable….
- ✓ Démographie, médecine : durée de la vie humaine, durée entre le déclenchement d'une maladie et la guérison, durée séparant deux naissances,…
- ✓ Economie, assurance : durée d'un épisode de chômage, durée de vie d'une entreprise, durée séparant deux sinistres, instant d'un défaut de paiement, durée avant la ruine, …
- ✓ Dans tous les domaines où l'on cherche à mesurer l'instant d'arrivée d'un évènement aléatoire (panne, mort, maladie, chômage,….)

I- Contexte et définitions

Spécificité des modèles de durées par rapport aux méthodes de statistiques classiques :

- ✓ Spécificité des données : échantillons de faible taille de durées parfois dépendantes, données incomplètes (censures, troncatures)
- ✓ Distributions non standard adaptées à la modélisation de phénomènes temporels et étude de fonctionnelles spécifiques
- ✓ Outils de modélisation non standard, adaptés à une observation séquentielle (stats d'ordre, processus ponctuels), différents suivant que l'évènement est renouvelable ou non.
- ✓ Variables explicatives exogènes pouvant dépendre du temps (région d'habitation, nombre d'enfants,..)

I- Contexte et définitions

Points de repères historiques :

- ✓ Premières études sur la mortalité en Angleterre au 17ème siècle par GRAUNT (1787) : définition des notions d'espérance de vie et d'espérance de vie résiduelle.
- ✓ Modélisation de la probabilité de décéder par GOMPERTZ (1825) *et* MAKEHAM (1860).
- ✓ WEIBULL (1951) propose un modèle paramétrique pour calculer la fiabilité d'un système non réparable. Il aborde notamment la présence de données tronquées ou censurées.
- ✓ KAPLAN et MEIER (1958) proposent d'utiliser dans le domaine médical un estimateur non paramétrique permettant d'intégrer les données censurées
- ✓ COX (1972) propose un modèle semi-paramétrique faisant intervenir des variables explicatives (exogènes).
- *LEE et CARTER (1992) proposent des* tables prospectives et modèles bi-dimensionnels « âge x année »

II- Les données

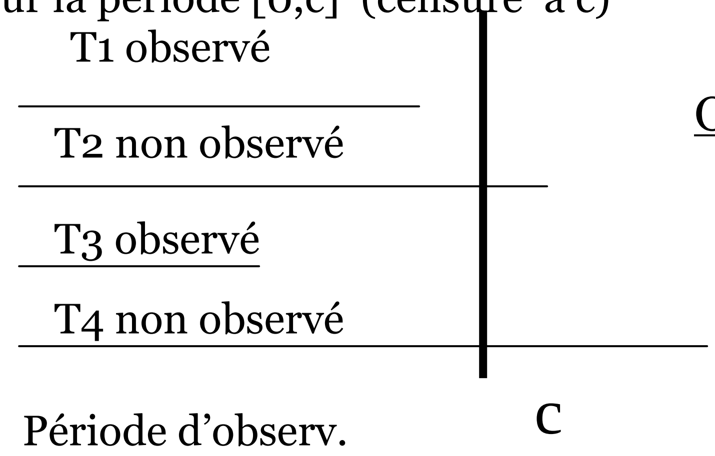
Spécificité des données :

- ✓ Difficulté à collecter un nombre suffisant d'observations (retours d'expériences) $\Rightarrow$ échantillons de faible taille

  Exemple : instants d'occurrence des pannes d'un système très fiable

- ✓ Difficulté à observer en continu le (les) phénomène $\Rightarrow$ données incomplètes

  Exemple : observation de l'instant d'occurrence du premier sinistre de 4 assurés sur la période [0,c] (censure à c)



Observation:

$$\boxed{X_i = T_i \wedge c ; \delta_i = 1_{T_i \le c}}$$

III- Les caractéristiques d'intérêt

A- indicateurs instantanés

- ✓ Disponibilité à l'instant t:

$$D(t) = P(\text{à } t, l'évènement\ n'est\ pas\ observé)$$

- ✓ Survie à l'instant t :

$$\mathrm{S}(t) = P(l'évènement\ ne\ se\ produit\ pas\ sur\ [0,t]) = P(T_1 > t)$$

- ✓ Maintenabilité à l'instant t :

$$\mathrm{M}(t) = P(\exists\ un\ instant\ de\ [0,t]\text{où } l'évènement\ n'est\ pas\ observé)$$

- ✓ Durée de vie moyenne restante à l'instant t (Mean Residual Life) :

$$\mathrm{r}(t) = E(T_1 - t|, T_1 > t)$$

III- Caractéristiques d'intérêt

A- indicateurs instantanés

- Taux de hasard (f densité de T1)

$h(t) = P(l'\text{évènement se produise à } t \text{ sachant qu'il ne s'est pas produit jusque là})$

$$\boxed{h(t) = \frac{d}{dt} P(t < T_1 \leq t + dt \mid T_1 > t) = \frac{f(t)}{S(t)}}$$

- Intensité de défaillance :

$\lambda(t) = P(l'\text{évènement se produise à } t \text{ sachant son histoire avant } t)$

$$\boxed{\lambda(t) = \frac{d}{dt} P(N(t+dt) - N(t) \geq 1 \mid \Psi_t)}$$

Où $N(t)$ est le nombre d'évènements produits jusqu'à l'instant t

Et Ψ_t est l'historique des évènements jusqu'à l'instant t

III- Caractéristiques d’intérêt
B- Indicateurs de tendance

- ✓ Espérance de vie (MTTF)

$$\boxed{MTTF = E(T_1) = \int_0^{+\infty} S(t)dt}$$

- ✓ MTTR (pour un séjour commençant à 0)= durée moyenne de séjour (ex: durée moyenne de séjour en panne, durée moyenne en chômage, ...)

$$MTTR = \int_0^{+\infty} (1 - M(t))dt$$

- ✓ MTBF=temps moyen entre deux évènements consécutifs

IIV- Evènement renouvelable/non renouvelable

Evènement non renouvelable (cours 2): évènement ne se produisant qu'une seule fois   Ex : durée de vie d'une ampoule électrique, mort, instant d'occurrence du 1° sinistre, …

$$\Rightarrow$$

- ✓ Un unique instant d'occurrence de l'évènement T
- ✓ Données de retour d'expériences incomplètes
- ✓ Concentration sur la modélisation de la loi de T caractérisée par son taux de hasard h ou sa survie S
- ✓ Techniques d'analyse de durées de vie classiques

IIV- Evènement renouvelable/non renouvelable

Evènement renouvelable (cours 6) : Evènement pouvant se produire plusieurs fois. Ex : pannes d'une centrale nucléaire, sinistres, maladie.

$$\Rightarrow$$

- ✓ Une séquence de temps d'occurrences $(T_1,....T_n)_{n\in N}$ généralement non i.i.d (ex fiab : l'état du système à t dépend de son histoire depuis l'origine jusqu'à t).
- ✓ La modélisation de la loi de la séquence des instants d'occurrences équivaut à celle du Processus de comptage $(N_t)_{t\geq 0}$ comptant à chaque instant le nombre d'occurrences de l'évènements survenues sur [0,t].
- ✓ Retours d'expériences sous-informés (troncature, censure, peu de réplication du processus d'occurrences)
- ✓ Concentration sur la modélisation de la loi de $(N_t)_{t\geq 0}$, caractérisée par son intensité de défaillance (la simplicité du modèle dépend de la capacité de mémorisation du système, ex fiab: liée au type de réparations subies et à leur qualité).
- ✓ Modèles de durées pour évènements renouvelables (Théorie des processus de comptage, ...)

Partie I : Modélisation d'évènements non renouvelables

Cours 2 : Généralités

I- Définition

II- Fonctionnelles étudiées

III- Propriétés du taux de hasard

IV- Familles de lois classiques

V- Les données de retour d'expérience

VI- Lois classiques pour la censure

VII- Simulation de durées de vies

I- Définition

Evènement non renouvelable=évènement ne pouvant se produire qu'une seule fois

- Grandeur d'intérêt : T=instant d'occurrence de l'évènement (par rapport à un instant de référence) est une durée de vie

$$\boxed{T = \text{durée de vie} = va \quad \geq 0}$$

- Objectif : modéliser la loi de T et estimer ses paramètres

- Méthodes : modèles de durée pour des évènements non renouvelables

II - Fonctionnelles étudiées

A- Caractéristiques principales

✓ Taux de hasard (ou taux de défaillance):

$$h(t) = \frac{d}{dt} P(t < T \leq t + dt / T > t) = \frac{f(t)}{1 - F(t)}$$

h positif, et $\int_0^{\infty} h(t)dt = \infty$

✓ Fonction de hasard : $H(t) = -Ln(1 - F(t)) = \int_0^t h(s)ds$

✓ Fonction de survie: (fiabilité) $S(t) = P(T > t) = 1 - F(t) = \exp(-\int_0^t h(s)ds)$

S monotone décroissante à valeurs dans [0,1]

✓ Fonction de répartition : $F(t) = P(T \leq t) = 1 - \exp(-\int_0^t h(s)ds)$

✓ Densité : $f(t) = F'(t) = h(t)\exp(-\int_0^t h(s)ds)$

Chacune de ces fonctionnelles determine entierement la loi de T

II- Fonctionnelles étudiées:

B- Caractéristiques secondaires

- ✓ Espérance de vie (MTTF, MTBF)

$$E(T) = \int_0^\infty t f(t)dt = \int_0^\infty S(t)dt \quad \text{si } \lim_{t\to\infty} tS(t) = 0$$

- ✓ Survie conditionnelle

$$S_u(t) = P(T > t + u / T > u) = \frac{S(t+u)}{S(u)} = \exp\left(-\int_u^{u+t} h(s)ds\right)$$

- ✓ Durée de vie moyenne restante

$$r(u) = E(T - u / T > u) = \frac{1}{S(u)} \int_u^\infty (t-u) f(t)dt$$

$$r(u) = \frac{1}{S(u)} \int_u^\infty S(t)dt \quad \text{si } \lim_{t\to\infty} tS(t) = 0$$

r determine entierement la loi de T

II - Fonctionnelles étudiées:
C- Cas d'une durée de vie discrète

- ✓ Densité $$\boxed{p_k = P(T = k) \quad k \geq 0}$$
- ✓ Fonction de survie $$\boxed{S(k) = P(T > k) = \sum_{i \geq k+1} p_i}$$
- ✓ Taux de hasard=taux de décès à l'âge k $$\boxed{h(k) = P(T = k / T > k-1) = \frac{p_k}{S(k-1)}}$$

- Relation entre survie et taux de hasard: taux de survie à k $$\boxed{S(k) = \prod_{i \leq k} (1 - h(i))}$$

III- Propriétés du taux de hasard

Quand on analyse des durées de vie, les 5 formes les plus usuelles de taux de hasard sont :

- ✓ Constant (durée de vie exponentielle)
- ✓ Croissant
- ✓ Décroissant
- ✓ En cloche
- ✓ En forme de baignoire

III- Propriétés du taux de hasard

- ✓ Un taux de hasard décroissant est caractéristique d'un système qui s'améliore (phase de déverminage ou de jeunesse, « burn in », mortalité infantile)

$$h \text{ décroit} \quad \overset{def}{\Longleftrightarrow} \quad \text{Loi de T DFR (decreasing failure rate)}$$

- ✓ Un taux de hasard constant est caractéristique d'un système qui ne vieillit pas (phase d'exploitation ou de vie utile)

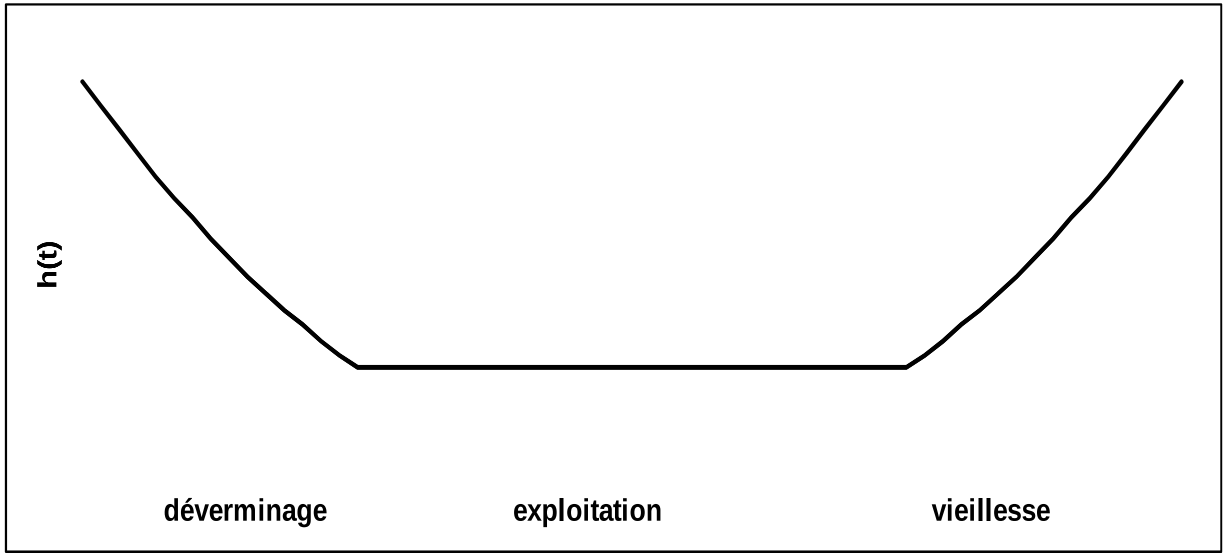
$$h \text{ constant} \quad \Leftrightarrow \quad \text{Loi de T exponentielle}$$

- ✓ Un taux de hasard croissant est caractéristique d'un système qui se détériore (phase d'obsolescence ou de vieillissement)

$$h \text{ croit} \quad \overset{def}{\Longleftrightarrow} \quad \text{Loi de T IFR (decreasing failure rate)}$$

III- Propriétés du taux de hasard

Comportement d'un système (homme, matériel) sur toute sa période de vie :



IV- Familles de lois classiques pour T

- ✓ d'une manière générale, toutes les distributions utilisées pour modéliser des variables positives (log-normale, Pareto, logistique, etc.) peuvent être utilisées dans des modèles de survie.

- ✓ La distribution de base des modèles de durée est la distribution exponentielle, et ses diverses généralisations (Weibull, Gamma)

- ✓ Le choix du modèle détermine en particulier la forme de la fonction de hasard ; on distinguera les modèles à fonction de hasard monotone des modèles permettant d'obtenir des fonctions de hasard en baignoire et en cloche;

- ✓ La situation de référence en assurance non vie est un taux de hasard constant ou croissant, en démographie en assurance vie et en fiabilité, le taux et en cloche ou monotone suivant la période étudiée

IV- Familles de lois classiques pour T

A- Loi exponentielle

- ✓ **Loi exponentielle** loi de la durée de vie d'un système durant sa phase d'exploitation

$T \sim E(\lambda), \lambda > 0 \Leftrightarrow$

$$\boxed{\begin{aligned} & f(t) = \lambda e^{-\lambda t} \\ & F(t) = 1 - e^{-\lambda t} \\ & h(t) = \lambda \\ & H(t) = \lambda t \\ & S(t) = S_u(t) = e^{-\lambda t} \qquad (*) \\ & E(T) = r(u) = \frac{1}{\lambda} \\ & V(T) = \frac{1}{\lambda^2} \end{aligned}}$$

Rq : (*) : le comportement su système après u ne dépend pas de ce qui est survenu jusqu'en u.

IV- Familles de lois classiques pour T

B- familles à taux monotones

- ✓ **Loi gamma** loi pouvant modéliser un système sur ses trois phases de vie; inadaptée pour la modélisation de la vie humaine (décroissance trop rapide)

$$T \sim \Gamma(\theta,\nu),\ \theta \text{ (forme)} > 0, \nu \text{ (échelle)} > 0 \Leftrightarrow$$

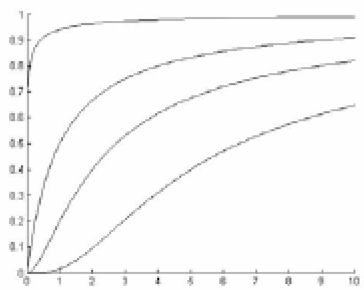
$$f(t) = \frac{\theta^{\nu}}{\Gamma(\nu)} t^{\nu-1} e^{-\theta t}$$

$$F(t) = \frac{1}{\Gamma(\nu)} \int_0^{\theta t} u^{\nu-1} e^{-u} du$$

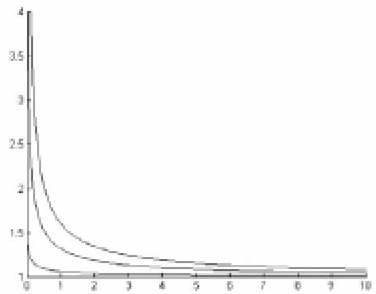
$$h(t) = \frac{f(t)}{1-F(t)}$$

$$E(T) = \frac{\nu}{\theta}$$

$$Var(T) = \frac{\nu}{\theta^2}$$



$r > 1$



$r < 1$

$$\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$$

IV- Familles de lois classiques pour T

B- familles à taux monotones

$0 < \nu < 1 \Leftrightarrow F$ DFR $\quad \lim_{h \to 0} h(t) = +\infty \,; \lim_{h \to \infty} h(t) = 1/\theta$

$\nu = 1 \Leftrightarrow T$ de loi exponentielle de paramètre θ

$\nu > 1 \Leftrightarrow F$ *I*FR $\quad \lim_{h \to 0} h(t) = 0 \,; \lim_{h \to \infty} h(t) = \theta$

IV- Familles de lois classiques pour T

B- familles à taux monotones

- ✓ La loi gamma est une généralisation de la loi exponentielle (qui correspond à $\nu=1$)
- ✓ Propriété: La somme de n variables exponentielles de même paramètre λ est une loi gamma de paramètres (n, λ).
- ✓ Lorsque ν est entier la loi gamma s'appelle Loi d'Erlang

IV- Familles de lois classiques pour T

B- familles à taux monotones

- ✓ **Loi de Weibull** Loi modélisant un système série (le système est défaillant dès lors qu'un de ses composants l'est), dans ses trois phases de vie

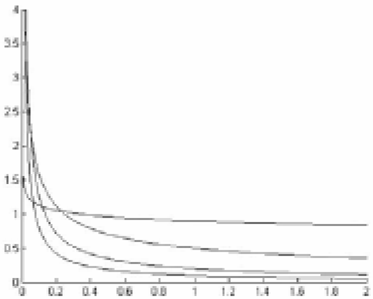
$$T \sim W(\beta,\alpha),\ \alpha\ (\text{échelle}) > 0,\ \beta\ (\text{forme}) > 0 \Leftrightarrow$$

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$$

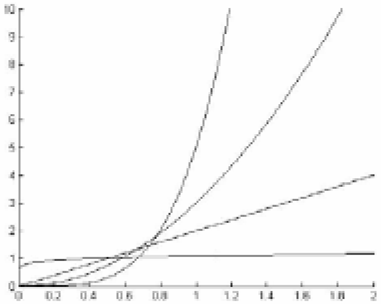
$$F(t) = 1 - e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$$

$$h(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}$$

$$E(T) = \alpha\Gamma\left(1+\frac{1}{\beta}\right)$$



α < 1



α > 1

IV- Familles de lois classiques pour T

B- familles à taux monotones

- ✓ $0 < \beta < 1 \Leftrightarrow$ F DFR $\lim_{h \to 0} h(t) = +\infty \,; \lim_{h \to \infty} h(t) = 0$
- ✓ $\beta = 1 \Leftrightarrow$ La loi de T est exponentielle $1/\alpha$
- ✓ $\beta > 1 \Leftrightarrow$ F IFR $\lim_{h \to 0} h(t) = 0 \,; \lim_{h \to \infty} h(t) = +\infty$

IV- Familles de lois classiques pour T

B- familles à taux monotones

- La loi de Weibull est une généralisation simple du modèle exponentiel, permettant d'obtenir des fonctions de hasard monotone.

- Lien avec la loi exponentielle : $\boxed{T \sim W(\beta,\alpha) \Leftrightarrow \left(\dfrac{T}{\alpha}\right)^{\beta} \sim E(1)}$

- Loi de RAYLEIGH : Lorsque $\alpha=1$ et $\beta=2$ ce modèle porte le nom de « modèle de RAYLEIGH» ; il est utilisé en physique pour modéliser la durée de vie de certaines particules.

- La loi de Weibull apparaît naturellement dans l'étude de la distribution limite du minimum d'un échantillon iid

IV- Familles de lois classiques pour T

B- familles à taux monotones

✓ **Loi de Gompert-Makeham** loi modélisant la mortalité humaine = modèle de référence pour la construction de tables de mortalité.

$$\boxed{h(t) = a + bc^t}$$

$$\boxed{S(t) = \exp(-at - \frac{b}{\ln(c)}\left(c^t - 1\right))}$$

a= taux de décès accidentel (indépendant de l'âge)

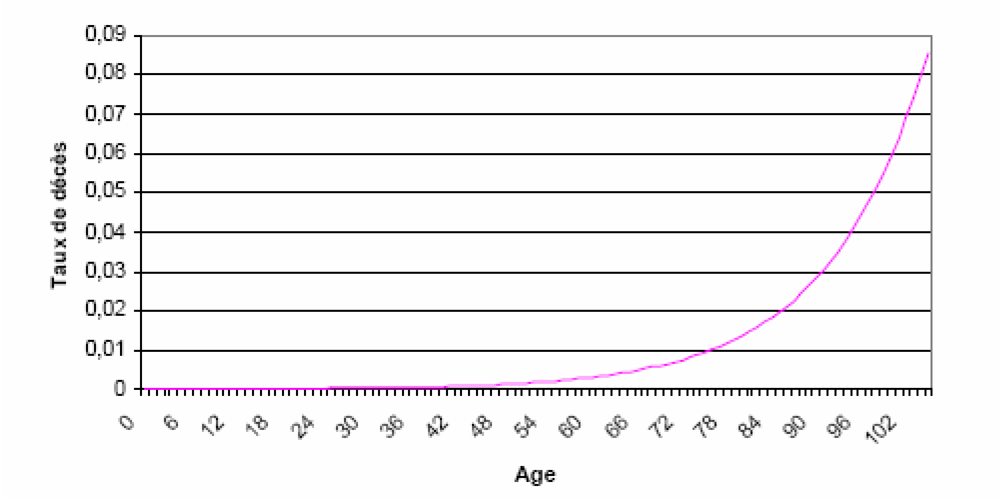
b*c^t=vieillissement exponentiel (si c > 1)

En mortalité humaine :

A	b	C
8,81E-06	3,83E-05	1,076207

IV- Familles de lois classiques pour T

B- familles à taux monotones



On peut noter graphiquement la croissance plus rapide du taux instantané de décès avec l'âge que dans le cas d'une loi de Weibull (cf. 3.2 ci-dessus), qui est en général mieux adapté à la mortalité humaine.

✓ Lorsque c=1, c'est une loi exponentielle de paramètre a+b

IV- Familles de lois classiques pour T

B- familles à taux monotones

- ✓ Mélange d'exponentielles

$$f(t) = \frac{p_1}{\theta_1} e^{-\frac{t}{\theta_1}} + \frac{p_2}{\theta_2} e^{-\frac{t}{\theta_2}} \qquad 0 < p_1 < 1, p_2 = 1 - p_1, \theta_1 > 0, \theta_2 > 0$$

$$S(t) = p_1 e^{-\frac{t}{\theta_1}} + p_2 e^{-\frac{t}{\theta_2}}$$

$$E(T) = p_1\theta_1 + p_2\theta_2$$

- ✓ Loi de Weibull généralisée

- ✓ Loi de Weibull exponentiée

IV- Familles de lois classiques pour T

C- familles à taux en cloches

- ✓ Loi log-normale de paramètre (μ,σ)

$$\boxed{F(t) = \Phi\left(\frac{\ln t - \mu}{\sigma}\right), \quad \sigma > 0, t > 0}$$

- ✓ Loi gaussienne inverse (μ,λ)

$$\boxed{f(t) = \sqrt{\frac{\lambda}{2\pi t^3}} \exp\left(-\frac{\lambda}{2\mu^2 t}(t-\mu)^2\right), \quad \lambda > 0, t \geq 0}$$

- ✓ Loi de Birnbaum et Saunders (α,β)

Adaptée à un mode de défaillance du à un phénomène de fatigue

$$F(t) = \Phi\left(\frac{1}{\alpha}\xi\left(\frac{t}{\beta}\right)\right), \quad \alpha > 0, \beta > 0, t > 0$$

$$\xi(t) = \sqrt{t} - \frac{1}{\sqrt{t}}$$

IV- Familles de lois classiques pour T

C- familles à taux en cloches

- ✓ Loi log-logistique de paramètre (α,β)

$$\boxed{F(t) = \frac{t^{\beta}}{\alpha^{\beta} + t^{\beta}},\ \alpha > 0, \beta > 1}$$

Ex: mortalité due au cancer après diagnostic ou traitement

IV- Familles de lois classiques pour T

D- familles à taux en baignoire

✓ Modèle quadratique

$$\boxed{h(t) = \alpha + \beta t + \gamma t^2 \quad -2(\alpha\gamma)^{1/2} \leq \beta < 0, \gamma > 0}$$

✓ Modèle de Weibull modifié

$$\boxed{h(t) = a(\alpha + \lambda t) t^{\alpha-1} e^{\lambda t} \quad a > 0, \alpha > 0, \lambda > 0}$$

✓ Modèle « exponential power »

$$\boxed{h(t) = \lambda\alpha(\lambda t)^{\alpha-1} e^{(\lambda t)^{\alpha}} \quad \alpha > 0, \lambda > 0}$$

V- Les données de retour d'expérience

- En vue de modéliser la loi de T, on observe les instants d'occurrences de n évènements identiques.

- En pratique :

  – n assez faible
  – L'information est généralement incomplète : on n'observe pas touts les évènements.

V- Les données de retour d'expérience

- ✓ Information complète : on observe tous les systèmes de leur mise en fonctionnement (naissance) jusqu'à leur instant d'occurrence :

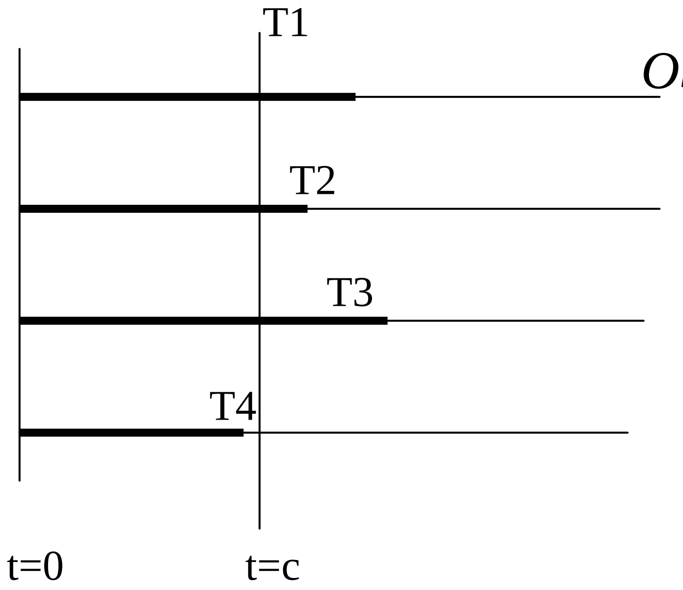
$$\boxed{Obs = (T_i)_{i=1,\ldots,n} \quad \text{i.i.d.}}$$

- ✓ Information incomplète : Au lieu d'observer des réalisations i.i.d. de durées T , on observe la réalisation de la variable T soumise à diverses perturbations, indépendantes ou non du phénomène étudié. Deux grandes familles de perturbations :
  - ✓ Censures : indépendantes du phénomène étudié
  - ✓ troncatures : dépendantes du phénomène étudié

V- Les données de retour d'expérience

A- Censure

- ✓ **Censure fixe gauche** : on observe les systèmes à partir d'une certaine date c, connue , jusqu'à l'occurrence de l'évènements :



$$Obs : \left(X_i, \delta_i\right)_{1\le i\le n} \,/ \begin{cases} X_i = T_i \vee c, \;\; \delta_i = \mathbb{1}_{T_i \ge c} \\ c \in \mathbb{R}, \quad (T_i)_{1\le i\le n} \quad \text{i.i.d.} \quad f \end{cases}$$

$$\boxed{\begin{array}{l} \text{Loi de } X : L(x) = F(x)\mathbb{1}_{c\le x},\, l(x) = f(x)\mathbb{1}_{c\le x} \\ \text{Loi de } \delta : \;\; \mathcal{B}(S(c)) \\ \text{Loi de } O : \;\; k(x,\delta) = \left(f(x)^{\delta} F(c)^{1-\delta} \right) \mathbb{1}_{c\le x} \end{array}}$$

Obs: {(T1,1);(T2,1),(T3,1);(c,0)}

1 censure, 3 observations de T

V- Les données de retour d'expérience
A- Censure

✓ *Démonstration densité des observations* :

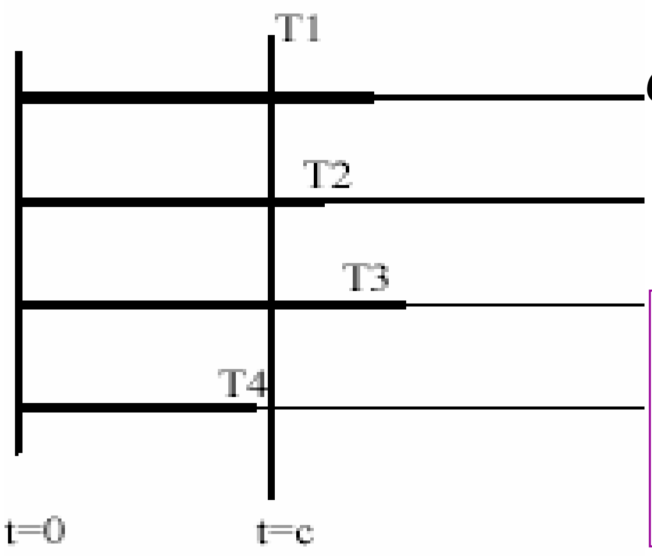
$$k(x,1)dx = P(x \leq X \leq x+dx, \delta = 1) = P(x \leq X \leq x+dx, c \leq T)$$
$$= P(x \leq T \leq x+dx) = f(x)\mathbb{1}_{x \geq c}dx$$
$$k(x,1) = f(x)\mathbb{1}_{x \geq c}$$

$$k(x,0)dx = P(x \leq X \leq x+dx, \delta = 0) = P(x \leq X \leq x+dx, T < c)$$
$$= P(x \leq c \leq x+dx, T < c) = F(c)\mathbb{1}_{x=c}dx$$
$$k(x,0) = F(c)\mathbb{1}_{x=c}$$

La distribution est continue par rapport à T et discrète par rapport à c.

V- Les données de retour d'expérience
A- Censure

- ✓ **Censure fixe droite (ou censure de type I):** on observe les systèmes de leur mise en fonctionnement jusqu'à une certaine date c connue



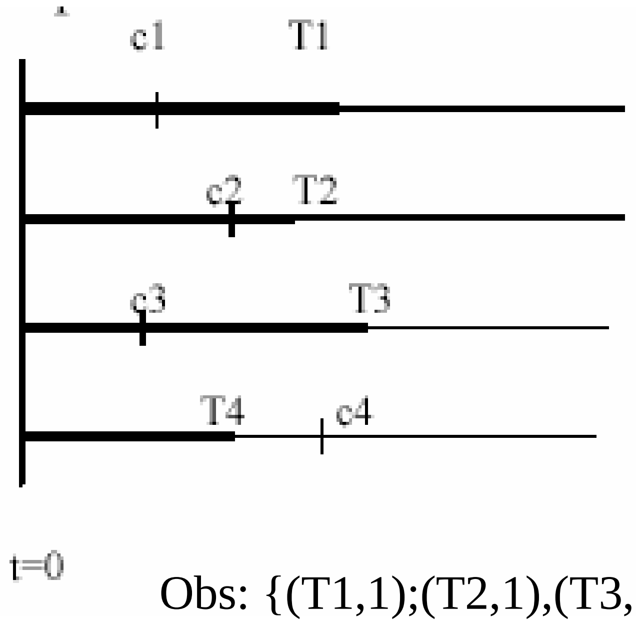
$$Obs : \left(X_i , \delta_i \right)_{1 \le i \le n} / \begin{cases} X_i = T_i \wedge c, \quad \delta_i = \mathbb{1}_{T_i \le c} \\ c \in \mathbb{R}, \quad (T_i)_{1 \le i \le n} \quad \text{i.i.d. } f \end{cases}$$

$$\boxed{\begin{array}{l} \text{Loi de } X : L(x) = 1 - S(x)\mathbb{1}_{c \ge x}, \; l(t) = f(x)\mathbb{1}_{c \ge x} \\ \text{Loi de } \delta : \; \mathcal{B}(F(c)) \\ \text{Loi de } O : \; \mathrm{k}(x,\delta) = \left(f(x)^{\delta} S(c)^{1-\delta} \right) \mathbb{1}_{c \ge x} \end{array}}$$

Obs: {(c,0);(c,0),(c,0);(T4,1)}
3 censures, 1 observation de T

V- Les données de retour d'expérience
A- Censure

✓ **Censure aléatoire gauche** : on observe les systèmes à partir d'une certaine date C aléatoire indépendante de T et de loi G jusqu'à l'évènement



$$Obs : \left(X_i, \delta_i\right)_{1\le i\le n} / \begin{cases} X_i = T_i \vee C_i, \quad \delta_i = \mathbb{1}_{T_i \ge C_i} \\ (T_i)_{1\le i\le n} \quad \text{i.i.d. } F, (C_i)_{1\le i\le n} \text{i.i.d. } G \\ T_i \perp C_i \end{cases}$$

$$\boxed{\begin{array}{l} \text{Loi de } X : L(t) = F(t)G(t); \quad l(t) = f(t)g(t) \\ \text{Loi de } \delta : \ \mathcal{B}(1-CP) \ \text{ avec } CP = P(T < C) \\ \text{Loi de } O : \ \mathrm{k}(x,\delta) = \left(f(x)G(x)\right)^{\delta}\left(g(x)F(x)\right)^{1-\delta} \end{array}}$$

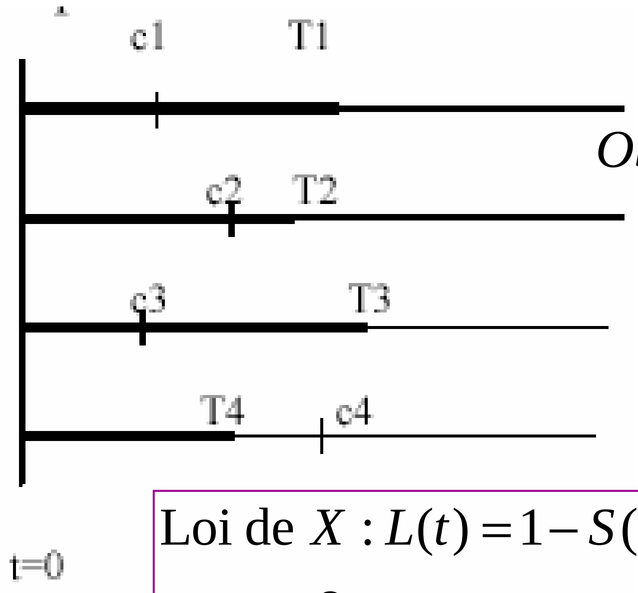
Obs: {(T1,1);(T2,1),(T3,1);(C4,0)}

1 censure, 3 observations de T

Rq : Généralisation du modèle de censure fixe(g remplace la mesure de Dirac au point c)

V- Les données de retour d'expérience
A- Censure

- ✓ **Censure de type III (aléatoire droite)** : on observe les systèmes de leur mise en fonctionnement jusqu'à une certaine date aléatoire C de loi G, indépendante de T.



$$Obs : \left(X_i, \delta_i\right)_{1\le i\le n} / \begin{cases} X_i = T_i \wedge C_i, \quad \delta_i = 1_{T_i \le C_i} \\ (T_i)_{1\le i\le n} \quad \text{i.i.d. } F, (C_i)_{1\le i\le n} \quad \text{i.i.d. } G \\ T_i \perp C_i \end{cases}$$

Obs: {(C1,0);(C2,0),(C3,0);(T4,1)}

3 censures, 1 observations de T

Loi de $X : L(t) = 1 - S(t)(1-G(t)); \quad l(t) = f(t)\left(1-G(t)\right) + g(t)S(t)$

Loi de $\delta :\ \mathcal{B}(1-CP)$ avec $CP = P(T > C)$

Loi de $O :\ \mathrm{k}(x,\delta) = \left(f(x)(1-G(x))\right)^{\delta}\left(g(x)S(x)\right)^{1-\delta}$

V- Les données de retour d'expérience
A- Censure

Démonstration de la densité des observations:

$$k(x,1)dx = P(x \le X \le x + dx, \delta = 1) = P(x \le X \le x + dx, C \ge T)$$
$$= P(x \le T \le x + dx, C \ge x) = f(x)(1 - G(x))dx$$
$$k(x,1) = f(x)(1 - G(x_i))$$

$$k(x,0)dx = P(x \le X \le x + dx, \delta = 0) = P(x \le X \le x + dx, T > x)$$
$$= P(x \le C \le x + dx)S(x)$$
$$k(x,0) = g(x)S(x)$$

V- Les données de retour d'expérience
A- Censure

- **Censure aléatoire de type II (censure au r° décès):** On observe jusqu'à l'occurrence du r° évènement (r° décès) . On est donc dans un modèle de censure de type I où $C = T_{(r)}$

(la r° statistique d'ordre de l'échantillon)

$$Obs : \left(X_i, \delta_i\right)_{1\le i\le n} \Big/ \begin{cases} X_i = T_i \wedge T_{(r)}, \quad \delta_i = 1_{T_i \le T_{(r)}} \\ \quad (T_i)_{1\le i\le n} \quad \text{i.i.d. } f \end{cases}$$

- Utile lorsque le phénomène est de faible fréquence.

V- Les données de retour d'expérience
A- Censures

- **Lois classiques pour la censure**

- Censure uniforme : $(C_i)_{1\leq i\leq n}\ i.i.d. \sim U[a,b]$
- Censure exponentielle : $(C_i)_{1\leq i\leq n}\ i.i.d. \sim E(\lambda)$
- Modèle de Koziol-Green:

$$\boxed{(C_i)_{1\leq i\leq n}\ i.i.d. \sim G \,/\, 1-G(t)=(1-F(t))^{\eta}}$$

- La valeur des paramètres règle le taux CP de censures dans l'échantillon.

  Ex : Dans un modèle de censure droite: $\boxed{CP = P(T > C)}$

V- Les données de retour d'expérience
A- Censures

✓ Exemple de calcul de CP

G suit une loi E(λ), F suit une loi E(1).

$$CP = P(T > C_\lambda) = \int_0^{+\infty} P(T > c)g(c)dc$$

$$= \lambda \int_0^{+\infty} e^{-(1+\lambda)c} dc = \frac{\lambda}{1+\lambda}$$

$$\boxed{\Rightarrow \lambda = \frac{CP}{1-CP}}$$

V- Les données de retour d'expérience
B- Troncature

- ✓ **Troncature gauche (resp. droite)** : on observe les systèmes à partir d'une certaine date Z connue ou aléatoire jusqu'à l'évènement (resp. de la naissance jusqu'à Z); avant (resp. après Z), on ne peut rien observer.

$$\boxed{\begin{array}{l} Obs:\left(T_i, Z_i\right)_{Z_i \le T_i} \\ Z_i \text{ i.i.d. de loi g}, \quad T_i \text{ i.i.d. de loi } f, T_i \perp Z_i \end{array}}$$

La distribution observée dans ce cas est la loi conditionnelle à l'événement $\{Z < T\}$.

$$\boxed{\begin{array}{l} \text{Loi : } k(t, z / T > Z) = \dfrac{f(t) g(z)}{P(T > Z)} \\ P(T > Z) = \int_0^\infty G(u) f(u) du \end{array}}$$

V- Les données de retour d'expérience
B- Troncature

NB : La phénomène de troncature est très différent de la censure :

- ✓ dans le cas de la troncature, on perd complètement l'information sur les observations en dehors de la plage d'observation: les systèmes ne sont pas observables, donc même pas répertoriés. La troncature élimine de l'étude une partie des systèmes.
- ✓ dans le cas de la censure, on a connaissance du fait qu'il existe une information, mais on ne connaît pas sa valeur précise, simplement le fait qu'elle excède un seuil ; dans le cas de la troncature on ne dispose pas de cette information.

V- Les données de retour d'expérience
B- Troncature

Exemples d'expériences de troncatures :

- ✓ migration informatique au cours de laquelle n'auraient été repris dans la nouvelle base que les sinistres encore en cours au moment de la bascule : les informations sur les sinistres de durée plus courte sont alors perdues.
- ✓ Contrat d'arrêt de travail avec une franchise : les arrêts de durée inférieure à la franchise ne sont pas observés, et on ne dispose donc sur eux d'aucune information.

V- Les données de retour d'expérience
C- Autres

- ✓ Censures par intervalles,
- ✓ troncatures par intervalles
- ✓ censure progressive
- ✓ Censures dépendant de la durée de vie

V- Les données de retour d'expérience
D- Identifiabilité

L'objectif dans les chapitres suivants est de modélisée la loi de T. La connaissance de la loi des observations incomplètes de T permet-elle d'identifier la loi de cette durée?

Pas toujours, car le fait de disposer de données incomplètes constitue une perte d'information

✓Dans le cas général, si la durée et la censure (ou la troncature) ne sont pas indépendantes, il n'est pas possible d'identifier la loi de T à partir des observations.

✓Dans le cas inverse, c'est souvent possible.

Mais….

V- Les données de retour d'expérience
D- Identifiabilité

La réponse n'est pas toujours positive car avoir des données incomplètes constitue une perte d'information. En effet, considérons le cas simple d'une durée T de support $\mathbb{R}^+$ et une censure fixe droite en un point t_0. Il ne nous sera pas possible d'identifier la partie de la distribution de T qui se trouve à droite de t_0 ; nous ne pourrons identifier que la probabilité que T dépasse t_0, mais nous ne pourrons pas préciser le nombre de modes, la durée de vie restante au-delà de t, pour tout $t > t_0$, l'épaisseur de la queue de distribution...

Dans le cas d'une troncature droite fixe en t_0, le problème est similaire. J'ai en fait encore moins d'information car je ne peux identifier que la loi de T conditionnelle à $T \leq t_0$. De plus, $P(T > t_0)$ ne pourra pas même être évalué.

V- Les données de retour d'expérience
D- Identifiabilité

Identifiabilité dans le cas de censures:

Soit sp(T) le support de la loi de T et $s(T) = \sup\{t \,/\, t \in sp(T)\}$

Proposition: Dans le cas d'une censure à droite aléatoire C, si T et C sont indépendantes et si $s(T) \leq s(C)$ alors la loi de T est identifiable à partir de la loi des observations .

V- Les données de retour d'expérience
D- Identifiabilité

Identifiabilité dans le cas de troncatures:

Proposition 1.5 *Supposons que*

$$P(Z > t) \neq 0 \Longrightarrow P(T > t) \neq 0.$$

Alors la loi du couple (Z,T) est identifiable.

V- Les données de retour d'expérience
D- Identifiabilité

Remarque: L'hypothèse d'indépendance entre durée et censure n'est pas toujours réaliste.

Ex : Lorsqu'une personne est soumise à plusieurs risques de décès (risques concurrents), ces risques sont tous dépendants de l'état général de santé (par exemple, la tension vasculaire).

Un risque particulier (par exemple, un infarctus) a de fortes chances d'être corrélé avec les autres (par exemple, un AVC) qui, si ils surviennent, le censurent éventuellement.

VI- Simulation de durées de vie : A- méthode de la fdr inverse

Simulation d'un échantillon de durées de vie pour T de fdr F :

1) F inversible :

Th : Si F inversible, U=F(T) suit une loi U[0,1]

- ✓ On simule $(u_1,\ldots,u_n) \sim U[0,1]$
- ✓ L'échantillon $(t_1,\ldots,t_n)$ est obtenu par $t_i = F^{-1}(u_i)$

2) F non inversible:

Th: Soit $I(u) = \inf\{t : F(t) \geq u\}$ alors $T = I(U)$

VI- Simulation de durées de vie

A- méthode de la fdr inverse

Ex : simulation de 10 durées de vie de loi E(1) sous R

```
>u=runif(10)
>t=-log(1-u,base=exp(1))
```

Ou

```
>t=rexp(1)
```

VI- Simulation de durées de vie

B- méthode de la fonction de hasard inverse

Simulation d'un échantillon de durées de vie lorsque T est de fonction de hasard donnée H :

1) H inversible :

th : Si H est inversible, V=H(T) suit une loi E(1)

- On simule $(u_1,\ldots,u_n) \sim U[0,1]$
- L'échantillon $(t_1,\ldots,t_n)$ est obtenu par $t_i = H^{-1}(-Ln(u_i))$

2) H non inversible:

th : Soit $K(v) = \inf\{t : H(t) \geq v\}$ alors $T = K(-Ln(U))$

VI- Simulation de durées de vie

C- échantillons censurés

- Simulation d'un échantillon de durées de vie censuré (droite)
  - On simule $(t_1,\ldots,t_n)$ de loi F
  - On simule $(c_1,\ldots,c_n)$ de loi G
  - L'échantillon est obtenu en calculant

$$\boxed{x_i = t_i \wedge c_i \qquad \delta_i = 1_{t_i \le c_i}}$$

- Simulation d'un échantillon de durées de vie censuré (droite) par C_λ de loi G exponentielle de paramètre λ, avec une proportion CP donnée de censure
  - On calcule λ / $CP = P(T > C_\lambda) = \int_0^{+\infty} P(T > c)g(c)dc$
  - On applique la procédure précédente